CIRCULAR No. 79 OCTOBER, 1929
UNITED STATES DEPARTMENT OF AGRICULTURE
WASHINGTON, D. C.

THE APPLICATION OF STATISTICAL METHODS TO SEED TESTING[1]

By G. N. Collins, *Principal Botanist, Biophysical Laboratory, Bureau of Plant Industry*

CONTENTS

INTRODUCTION

During the last 20 years there has been a continued increase in the use of statistical methods. In practically all the branches of science that involve quantitative data biometrical or statistical methods have been found helpful. For the seed analyst some of these methods would seem indispensable, and the economic importance of seed testing justifies a careful study of the problems involved by a competent statistician. To determine the best and most accurate methods is a task for the trained mathematician, and although the development of the formulas may be highly technical analysts may look forward to a series of procedures that will be simple in their application.

It is the aim of this circular to indicate possible applications of some of the more simple formulas that have found use in other branches of science.

The most obvious application of statistical methods to the problems of seed testing is in providing measures of accuracy for the various determinations. Whenever a quantitative determination is recorded, whether it be a percentage of germination, the number of dodder seeds in a sample of clover, a temperature, or the length of a stick, some measure of accuracy is needed if the value recorded is to have definite meaning. When none is given the reader must supply it. This is often done unconsciously, but it is done nevertheless. If it is stated that a given plant is 3 feet high it is not assumed that the plant is exactly 3 feet high. Most readers would probably picture a plant somewhere between 2½ and 3½ feet high. If inches are given, the

[1] Paper read before the Association of Official Seed Analysts of North America at New York, N. Y., January 2, 1929.

55471—29——1

reader has a right to assume an accuracy of the order of 1 inch, though this is not always warranted.

Everyone knows that if a measurement is repeated it will not always give exactly the same result, and consciously or unconsciously some latitude is allowed. Those who refuse to attach some measure of accuracy to a determination should realize that for the omission they are compelling the reader to substitute a guess.

Any series of determinations based on the same material will be grouped around a mean value from which the individual determinations will depart more or less. If the values are widely scattered it is known that the accuracy is low and that a large error must be expected in measurements of this kind. Some measure of this error is needed in order that the reliability of such measurements may be estimated.

MEASURES OF VARIATION

Nothing definite can be learned from the limits of the scatter, or range, for this is not a fixed value, but continues to increase as the number of observations is increased.

The departures from the mean value may be averaged by summing the amounts by which each value departs from the mean and dividing by the number of observations. This value is termed the average deviation. As the number of observations increases the average deviation approaches a fixed value, and can be and has been used as a measure of accuracy.

Although an average deviation may be used as a measure of accuracy, a slightly different value has been found much more serviceable. It is called the standard deviation. A standard deviation is calculated by squaring the departures from the mean before they are averaged and then extracting the square root of the average. This gives a value about 20 per cent greater than the average deviation. It differs from an average deviation by giving an increased weight to wide departures.

There are many who become discouraged as soon as a standard deviation is mentioned. It has a rather unfamiliar sound, but if it is thought of as a slight modification of the average of individual departures from the mean it may appear less formidable.

Just as the height of two groups is compared by comparing the average heights, so in comparing the variation of one group with that of another it is possible to use the average variation or standard deviation. The standard deviation has many properties that render it exceedingly useful in a great variety of ways.

Before the uses of the standard deviation are discussed, one more term should be introduced. This is "variance," a term applied to the square of the standard deviation. It will be recalled that in calculating the standard deviation of a normal distribution the squared departures were averaged. This mean of the squared deviations is the variance, and for many calculations it is simpler to use it in this form without extracting the square root.

The formula for variance and standard deviation is—

$$\sigma=\sqrt{\Sigma d^2/N} \qquad (1)$$

where σ^2 = variance, σ = standard deviation
d = departure from mean.
N = number, Σ = summation.

The application of formula 1 may be illustrated by a series of 10 germination tests (Table 1):

TABLE 1.—*Series of 10 germination tests, illustrating the application of formula 1*

Test No.	Percentage of germination	d	d^2	Test No.	Percentage of germination	d	d^2
1	80	0	0	7	79	−1	1
2	82	+2	4	8	78	−2	4
3	77	−3	9	9	82	+2	4
4	81	+1	1	10	84	+4	16
5	79	−1	1				
6	78	−2	4	Mean	80		$44=\Sigma d^2$

The formula gives—
$\Sigma d^2/N=44/10=4.4=\sigma^2$ or variance. $\sqrt{\sigma^2}=\sqrt{4.4}=2.1=\sigma$ or standard deviation.

The standard deviation or variance may be used not only to measure the accuracy of single determinations but also to measure the accuracy of means or averages. If the variance of individual determinations is known, the variance of the mean of any number of such determinations will be the individual variance divided by that number; or, in dealing with standard deviations, the standard deviation of the mean will be the standard deviation of individual determinations divided by the square root of the number.

To speak of the variance or standard deviation of a single mean. really implies that if there were a series of such means they would be distributed about a general mean with the stated standard deviation.

In most experimental work an entire population from which to calculate a standard deviation is not to be had, but the worker must take as the best estimate of the standard deviation one derived from a comparatively small number of observations. If the numbers are few this estimated standard deviation will be too small, and to correct for the difference in deriving the variance or the standard deviation of a mean from the individual variance it is necessary to divide by the number minus 1 instead of by the number.

The formula for the standard deviation of a mean is—

$$\sigma_M=\sqrt{\sigma^2/N-1} \qquad (2)$$

where σ_M = standard deviation of the mean,
σ = standard deviation of individual values,
N = number of individual values.

As an example, assume a series of 10 germination tests. The mean of the 10 tests being 75 per cent and the variance of the individual tests being 18 per cent, the variance of the mean is then $18/9=2$ per cent, and the standard deviation is $\sqrt{2}=1.4$ per cent. The mean may be written 75 ± 1.4 per cent. This indicates that if the tests were repeated until there was a series of means each based on 10 tests these means might be expected to have a standard deviation of 1.4 per cent. The standard deviation of a mean, or the best estimate of it that can be made, is sometimes called the standard error. More frequently 0.6745 of the standard deviation is termed the probable error. In American publications the plus-or-minus sign (±) placed before a value indicates a probable error.

The reason for taking this particular fraction of the standard deviation as a measure of accuracy is that on an average it will include one-half of the deviations. As a measure of tolerance the probable error has little to recommend it except usage. No one would place confidence in a result that will be right only one-half or three-quarters of the time.

It has become a rather general practice to take three times the probable error as the limit of significance. Since the probable error has always to be multiplied by 3 or some other number before significance can be estimated, the use of the probable error would seem to have no advantage over the standard deviation, and there is the disadvantage that probable errors must be converted back into standard deviations before probability tables can be used.

Engledow and Yule go so far as to say that "Statement of the probable error in modern work is an unmitigated nuisance, and the investigator is recommended to confine himself to the standard error and accustom himself to thinking in terms of it.[2]

If seed analysts are not already committed to the use of probable errors, the use of the standard deviation or standard error is recommended.

TYPES OF DISTRIBUTION

Before taking up the uses of the standard deviation it will be necessary to consider the kinds of distributions for which measures of accuracy are needed.

From a statistical standpoint, the determinations which the seed analyst is called upon to make fall into three classes, each characterized by a particular frequency distribution. These three distributions are (1) the normal frequency, (2) the binomial distribution, and (3) the Poisson series.

NORMAL FREQUENCY

To the normal curve is referred all determinations such as weights and linear measurements that may have any value and that occur with approximately equal frequency above and below the mean value.

The mathematical requirements for a normal distribution are rather rigid, and few biological data conform completely. Fortunately, however, the methods worked out for the normal distribution may be applied without appreciable error to distributions not strictly normal.

Most of the applications of statistical methods to the problems of seed testing involve the use of tables of the normal or Gaussian distribution. To use these tables intelligently it is desirable to understand something of how they are constructed.

A frequency distribution assumes a large number of measurements thrown into groups of a uniform range. Assume 1,000 measurements of length, for example, having a mean of 40 centimeters and a class range of 5 centimeters. The first group above the mean would include all measurements between 40 and 45 centimeters, the second

---

[2] ENGLEDOW, F. L., and YULE, G. U. THE PRINCIPLES AND PRACTICE OF YIELD TRIALS. 78 p., illus. London. 1926. (From Empire Cotton Growing Rev. 3: 112–146, 235–267, illus. 1926.)

group those values falling between 45 and 50 centimeters, etc. Below the mean there would be a corresponding series of descending values. The groups near the mean will contain large numbers, and the number in each successive group will be less as the distance from the mean increases. If the standard deviation is small, a very large proportion of values will fall in the central groups, and the measurements will extend but a short distance from the mean. Assuming the standard deviation to be 10 centimeters, the values are plotted in Figure 1.

The first step in plotting the values is to transpose the scale into terms of the standard deviation. This is done by expressing the values as departures from the mean and dividing by the standard deviation. Thus 35 centimeters departs from the mean −5 centimeters. This divided by 10, the standard deviation, gives −0.5. The class limits in terms of the standard deviation are given in the second line below the base, labeled σ. If the 1,000 measurements are distributed in perfect agreement with a normal frequency there should be one individual below 10 centimeters, five individuals between 10 and 15 centimeters, etc., as in the diagram. The sum of these numbers is 1,000.

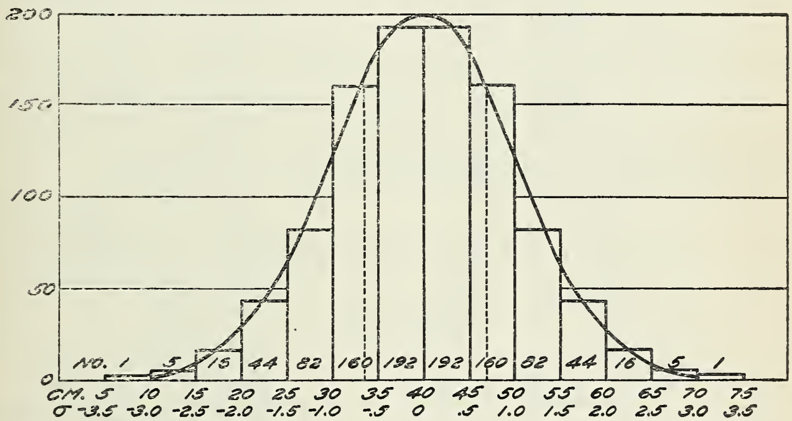

FIGURE 1.—Normal frequency polygon and curve. Class range 5 centimeters, standard deviation 10 centimeters. Total population, 1,000

It is easily understood that if the measurements had been grouped into smaller classes, say of 1 centimeter instead of 5, the steps in the polygon would have been less abrupt and the boundary would more closely approximate the smoothed curve that has been drawn. This is the curve of normal frequency, and if properly drawn it incloses the same area as the polygon. Theoretically, the curve does not touch the base, but extends to infinity in either direction. However, the area of the curve beyond the limits drawn is so small that it may usually be neglected or included in the last groups, as has been done in the diagram. The area in the curve beyond the limits of the diagram is less than one-fourth of one unit.

The two dotted lines of the diagram are so placed as to include between them one-half the total area. The lines are equidistant from the mean, and this distance is 0.6745σ, or about two-thirds of the standard deviation, in this instance 6.745 centimeters. It is now apparent that if from this population individuals are taken at random, one-half of them will depart from the mean by less than 0.6745σ and one-half will depart by more than that amount. This departure of 0.6745σ is called the probable error.

These relations hold good for all normal distributions, regardless of the units in which the measures are recorded and whether the population is uniform or variable. In Figure 2 a more uniform population is plotted to the same scale. Here the standard deviation is taken as 5 centimeters instead of 10 centimeters. As before, there are 1,000 units in the entire polygon. The two classes adjacent to the mean each contain 352 individuals, as compared to 160 in Figure 1, but if the classes are taken in terms of the standard deviation the number in corresponding groups is the same. Thus, in Figure 1 a distance from the mean equal to the standard deviation includes two groups which together contain 352 individuals. By expressing the departures from the mean in terms of the standard deviation all normal frequencies may be compared.

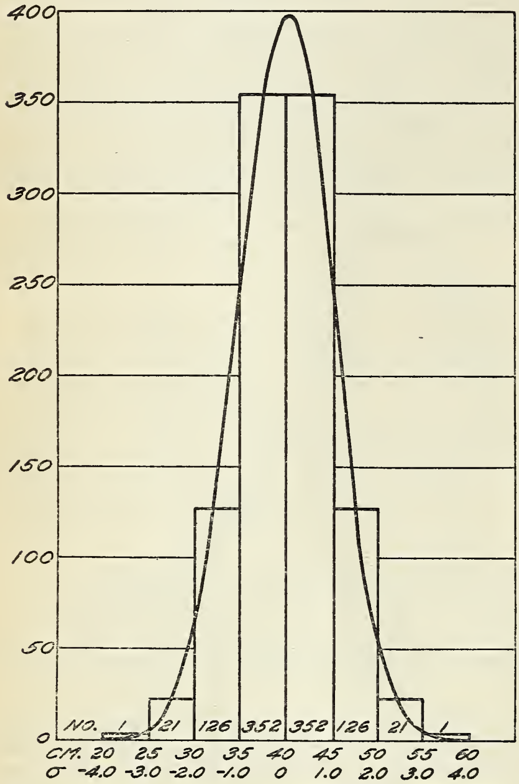

FIGURE 2.—Normal frequency polygon and curve. Class range 5 centimeters, standard deviation 5 centimeters. Total population, 1,000

Normal frequency tables are designed to give the ratio of the areas of the two portions into which the normal curve is divided by a line erected at any point on the base. Since each individual occupies a unit of area, the number of individuals may be substituted for the area.

The distances along the base measured from the mean or center of the curve are the abscissas, and perpendicular lines from the base to the boundary of the curve are the ordinates. The relations of area, abscissa, and ordinate have been tabulated in a number of ways. For present purposes the relation between area and abscissa is all that need be considered. Perhaps the most widely used tables are those of Sheppard,[3] which have been republished by Pearson [4] as his Table 2. In these tables the argument or primary division given in the first column is the abscissa, or departure from the mean, measured

[3] SHEPPARD, W. F. NEW TABLES OF THE PROBABILITY INTEGRAL. Biometrika 2: [174]-190. 1902.
[4] PEARSON, K. TABLES FOR STATISTICIANS AND BIOMETRICIANS. Ed. 2, pt. 1, 143 p., illus. London [1924].

in terms of the standard deviation, and is labeled "x." In the second column, labeled "1/2 $(1+a)$," is given the corresponding area of the larger portion of the entire curve or population cut off by the ordinate at point "x." The area of the entire curve is taken as one. The tabulated entries, therefore, are fractions beginning at 0.5 or one-half the population for $x=o$.

C. B. Davenport,[5] in his Table 4, gives a slightly different arrangement. The abscissas or departures from the mean are labeled x/σ and the values are given to three places instead of two. The area of only one half of the curve is considered, and the total area is taken as 100,000. To make the values equal those of Pearson's table it is necessary to move the decimal point five places to the left and to add 0.5.

The Kelley-Wood tables[6] have the reverse arrangement. The argument is the portion of the curve between the mean and the point x and is labeled "I." The corresponding values of x are given in the second column. In addition, under the columns headed "p" and "q" are given the larger and smaller portions of the entire curve; p is thus $0.5+I$ and $q=0.5-I$.

This arrangement is convenient whenever the portion of the population is known and the desire is to find the corresponding departure from the mean. As an example, suppose the object is to find the tolerance that would make the degree of certainty 1 in 100, or 0.01. Opposite $q=0.01$ or $p=0.99$ the value of x is given very exactly as 2.326348. In Pearson's or Davenport's tables 0.99 of the area in the values under 1/2 $(1+a)$ can not be located exactly without interpolation. It can be seen, however, that x will lie somewhere between 2.32 and 2.33.

Equivalent terms in the three tables are given in Table 2.

TABLE 2.—*Comparison of equivalent terms in the tables of Kelley-Wood, Pearson, and Davenport*

Term	Kelley-Wood	Pearson	Davenport
Departure from the mean in terms of the standard deviation	X	X	x/σ
Area between ordinates at the mean and at X	I	$[1/2(1+a)]-0.5$	$\frac{1/2a}{1,000}$
Area in larger portion of entire curve	p	$[1/2\ (1+a)]$	$\frac{1/2a}{1,000}+0.5$
Area in smaller portion of entire curve	q	$[1-1/2\ (1+a)]$	$0.5-\frac{1/2a}{1,000}$

If it is remembered that area and population are interchangeable and that departures from the mean are expressed in terms of the standard deviation, there should be little difficulty in using tables of the normal frequency.

In the applications that follow where tables of the normal frequency are used to determine probabilities the smaller portion of the entire area is the value used. This is the Kelley-Wood q or $1-$Pearson's ½ $(1+a)$, where the entire area or population is taken as unity. Thus

[5] DAVENPORT, C. B. STATISTICAL METHODS, WITH SPECIAL REFERENCE TO BIOLOGICAL VARIATION. Rev. Ed. 3. 225 p., illus. New York and London. 1924.
[6] KELLEY, T. L. STATISTICAL METHOD. 390 p., illus. New York. 1923.

at the mean or zero departure the value is 0.5, i. e., one-half the population will have the value of the mean or less than the mean. At 2, or twice the standard deviation above the mean, the value is 0.023; that is, 97.7 per cent of the population is below this value and but 2.3 per cent will exceed the mean by more than twice the standard deviation. Since the distribution is symmetrical, there will also be 2.3 per cent that fall below the mean by more than twice the standard deviation. In using probabilities as a measure of certainty it should be noted whether the probability is to include deviations in one or both directions. It is assumed that in most questions of tolerance it is desired to measure the probability in but one direction. But in estimating the significance of a difference between two samples, variations in both directions should be taken.

BINOMIAL DISTRIBUTION

The second type of distribution, the binomial, is one in which the values must fall into a limited number of classes, in numbers corresponding to the coefficients of an expanded binomial. As an example, if three coins are repeatedly tossed and the number of heads recorded there are but four possible values—no heads, one head, two heads, or three heads. No intermediate values are possible. The distribution of the numbers into these four classes will be in the ratio of 1/8:3/8:3/8:1/8. These numbers are the coefficients of $0.5p+0.5q$ raised to the third power, p being the expectation of getting a head and q the expectation of getting a tail.

A germination test involving 100 seeds corresponds to a case where 100 coins are tossed, but in this case the expectation of throwing a head will not be 0.5 unless the germination is 50 per cent. If the germination is 80 per cent the expected distribution would be given by $0.8p+0.2q$ raised to the hundredth power, a laborious process that fortunately does not have to be performed.

So far as the seed analyst is concerned, binomial distributions call for no special treatment. The values to be dealt with are percentages or means, and these will be normally distributed, making possible the use of the tables of normal distribution. The only difference in procedure is that here a short cut for deriving the standard deviation of a ratio or percentage may be utilized.

The formula for the standard deviation of a percentage is—

$$\sigma\% = \sqrt{pq/N} \qquad (3)$$

where $\sigma\%^2$ = variance of percentage,
$\sigma\%$ = standard deviation of percentage,
p = the percentage,
$q = 100 - p$,
N = number.

If N is small, $N-1$ should be substituted for N. It will be noted that this formula differs from that for the standard deviation of a mean by the substitution of pq for the individual variance.

POISSON SERIES

The Poisson series, the third kind of distribution, could have no better illustration than is given by the distribution of the number of dodder seeds in successive small samples of clover seed. The values are restricted to whole numbers, and if the mean number is small the distribution is far from symmetrical.

The departure of a Poisson series from a symmetrical distribution is evident if the frequency of numbers from samples that should contain one dodder seed is considered. There can be but one class below the mean, that is, samples with no dodder seeds, while above the mean there will be samples with 2, 3, 4 or more seeds. In fact, samples with no seeds will be much more frequent than samples with 1, the mean number.

Special probability tables giving the expected distribution of samples with means from 1 to 30 have been published, and from these the probability of departures from the mean of any given magnitude can be read directly.

When the mean or expected number is above 30 the distribution is practically symmetrical, and tables of the normal distribution may be used. In a Poisson series the variance equals the mean number. For example, if a lot of seed has dodder at the rate of 36 dodder seeds per 100 grams, the number of dodder seeds in successive samples of 100 grams will have a variance of 36, or a standard deviation of 6 seeds.

APPLICATION OF THE STANDARD DEVIATION

In the following illustrations of the standard deviation as a mesaure of the errors in seed testing, no account has been taken of the personal equation or individual differences in judgment.

By adopting rigid and specific rules for testing, individual variation can be greatly reduced, but it can never be eliminated entirely. The omission of this type of variation from present consideration should not be interpreted as indicating that individual differences in judgment are not a fit subject for statistical analysis. Determinations made by different operators may be treated in the same manner as different determinations by the same operator.

In what follows, however, it is assumed that all tests are made by the same individual.

With the exception of tests for noxious seeds when the number in the sample is small, all measures of accuracy and degrees of certainty that the seed analyst will encounter may be interpreted by means of the table of normal probabilities. Examples may be taken with equal propriety from either purity or germination tests.

As the first example, take the standard deviation as a measure of accuracy in weighings of impurities.

Suppose it is known that the weights of impurities have a standard deviation of 1 milligram. That is, repeated weighings of the same lot of impurities are distributed around a mean with a standard deviation of 1 milligram. Suppose, now, the purity of a stock of seed is said to be such that in the sample tested there should be less than 10 milligrams of impurities, and the sample is found to have 13 milligrams of impurities. This is a difference or departure of 3 milligrams and is three times the standard deviation. Opposite 3 in the probability tables is the value 0.9987, indicating that 99.87 per cent of the weighings of samples having a mean of 10 milligrams of impurities would show 13 milligrams or less. There is then a probability of 0.0013, or 1 in about 750, that the departure is the result of errors in weighing.

This determination tells nothing regarding errors from other sources than weighing, and without further knowledge it would be entirely unwarranted to assume that the impurities in the original lot of seed

55471—29——2

were in excess of the stipulated quantity. It does indicate, however, that errors of weighing will not account for the difference.

In using the standard deviation as a measure of the limits of random sampling, great care should be exercised that the variant measured has not been subject to sources of variation that were not taken into consideration when the standard deviation was determined. In the above illustration the standard deviation was determined from repeated weighings of the impurities from the same sample, whereas the 13 milligrams of impurities in the example was subject to errors of sampling in addition to the errors of weighing. In this example the distinction is obvious, but much of the criticism against the use of statistical methods is occasioned by overlooking this distinction in less-obvious instances.

Mistakes in the direction of too great leniency are less difficult to guard against. Thus if the sample gave 11 milligrams of impurities, a standard deviation of 1 milligram indicates that although the sample shows impurities in excess of the quantity stipulated, there is about one chance in three that the departure is due to an error of weighing, and there is little occasion to look farther for an explanation of the difference.

Another use of the standard deviation is in determining whether two samples may represent random variations from the same population or must have been drawn from different populations. Another way of stating this problem is to ask whether the difference between the two samples is significant. The probability that shall be taken as significant is a matter of individual preference, but the standard deviation will tell what the probability is, which may or may not be accepted.

As an illustration, suppose two purity tests have been made from what is assumed to be the same lot of seed. Suppose it is known that the weighings of impurities have a standard deviation of 2 milligrams. That is, repeated weighings of the same lot of impurities are distributed about a mean with a standard deviation of 2 milligrams. If the two tests gave 15 and 24 milligrams, could this difference of 9 milligrams be ascribed to random errors of weighing?

The variance of a difference between two determinations is the sum of the two individual variances. The formula for the variance and standard deviation of a difference is—

$$\sigma\Delta = \sqrt{\sigma_a^2 + \sigma_b^2} \qquad (4)$$

where $\sigma\Delta^2$ = variance of difference between a and b,
$\sigma\Delta$ = standard deviation of difference,
σ_a^2 = variance of a,
σ_b^2 = variance of b.

In this instance $2^2+2^2=8$ as the variance of the difference. The square root of 8, or 2.8, is the standard deviation of the difference. The difference between the two weighings is 9, or 3.2, times the standard deviation of the difference. The table gives the probability of 3.2 times the standard deviation as 0.0014. Since either value might have been the greater, take twice this value, or 0.0028—about 1 in 350.

It is evident that it can not be concluded from these two tests that the samples were not drawn from the same bulk. To test this point, the standard deviation of a series of samples drawn from the same bulk must be determined. With the new standard deviation derived from repeated samplings the procedure is the same.

The most frequent use of the normal distribution and accompanying tables is in connection with differences in mean values.

In dealing with means, a normal distribution may be assumed, although the individual values are distributed in accordance with the binomial or Poisson series. In germination tests the individual measurements can take but two values, dead or alive, and all percentages of germination are mean values. With the number of seeds used in any practical germination test the percentages of germination will be distributed in close agreement with a normal frequency, and the table of probabilities may be used to decide whether a difference in the percentage of germination of two samples may be ascribed to chance fluctuations from a common mean, or whether some other explanation must be sought.

As an illustration, assume that one test of 200 seeds gives 75 per cent germination. Another test of 100 seeds gives 66 per cent germination. Is there a significant difference between the two tests? The mean of both tests is $\frac{(75\times 200)+(66\times 100)}{300}=72$ per cent; $p=72$, $q=28$, $pq=2{,}016$.

By formula 3 variance first test $=2{,}016/200=10.08$, variance second test $=2{,}016/100=20.16$.

By formula 4 variance of difference $=10.08+20.16=30.24$. Standard deviation of difference $=\sqrt{30.24}=5.5$, $\Delta=75-66=9$, $\Delta/\sigma\Delta=9/5.5=1.64$. The probability table for 1.64σ gives 0.0505. Since either test might have been the larger the probability is $2\times 0.505=0.101$, or 1 in about 10. There is thus slight reason to go beyond the error of sampling for an explanation of the difference.

Applications of the standard deviation could be multiplied indefinitely, but it is hoped that the foregoing examples are sufficient to enable the reader to use it in other related problems. Although the calculations are extremely simple, they may seem laborious as a routine procedure. For any particular application that should be adopted as standard procedure most of the calculation could be eliminated by the construction of special tables. For example, it would be the work of but a few hours to prepare a table that would give at once the variance and standard deviation for all combinations of percentage and number of seeds.

DEGREE OF CERTAINTY

Most of the problems that confront the seed analyst can be summarized into the single question: Does an observed value differ from some other value by an amount too great to be ascribed to unavoidable error?

This question can not be answered with complete certainty, because there are no absolute limits to the errors of sampling. The most that can be achieved is such a high degree of certainty as will make the reporting of a mistake very improbable. The degree of this improbability is not a matter of statistics, but must be determined by considerations of personal responsibility.

What statistical methods can do in this connection is to enable the operator to maintain any particular standard that is decided upon.

For the present this standard may be termed the degree of certainty and expressed as a probability. For example, if it is decided that satisfactory protection is achieved with a procedure by which the chances against error are 99 to 1, that degree of certainty would be indicated by the probability expression 0.01, whereas 0.001 would indicate a degree of certainty of not more than 1 mistake in 1,000 tests.

The degree of certainty may vary with the use to be made of the findings, but not with the nature of the data or methods of calculation. It should not be confused with such measures of accuracy as standard error and probable error. If the percentage of germination is given as the result of a test, the observed value may be accompanied by an expression of the standard error as a measure of the accuracy of the determination. But if one is to go on record with a positive statement that the true germination is less than a certain percentage, some degree of certainty should be adopted. The standard error is a technical descriptive term, whereas degree of certainty like "margin of safety" indicates the element of risk.

TOLERANCE

To fix the limits of tolerance in any definite way it would seem desirable to recognize that there are two factors, more or less interchangeable, to be considered. These factors are (1) the magnitude of the errors involved in making the determinations and (2) the degree of certainty that is to be adopted.

The standard deviation gives a measure of the errors that may be expected as a result of errors of sampling, etc., but the degree of certainty is a matter of choice. No matter what tolerance is allowed, there is always some possibility that a variation of testing will exceed that limit. It may be urged that tests should be repeated. Repetition reduces the error but does not eliminate it, and it must be admitted that since in practice there is a limit to the number of repetitions, or, in other words, to the size of the sample tested, there will always be a certain, though perhaps small, proportion of cases where injustice is done. Statistical analysis makes it possible to fix that proportion rather definitely. Take, for example, a stated germination of 80 per cent and a standard deviation of 3 per cent. If a tolerance of twice the standard deviation, or 6 per cent, is allowed, there will be on the average 2.3 per cent of those cases having a true germination of 80 per cent that will be reported as falling below the tolerance and about the same percentage of cases having a germination of 74 per cent that will be reported as 80 per cent or above. If this percentage of mistakes is considered too large there are two ways of reducing it: (1) By reducing the standard deviation, that is, by increasing the size of the sample or perfecting the method of testing in some other way, or (2) by increasing the tolerance.

By increasing the tolerance from 6 to 7 per cent the percentage of errors of border-line cases will be reduced to 2 per cent—1 per cent above and 1 per cent below.

Published tables have eliminated all calculations from the determination of these percentages. If the tolerance is expressed in terms of the standard deviation, the probability or percentage of cases that will fall outside the tolerance may be read directly from the tables as a single value.

TOLERANCE IN PURITY TESTS

The unavoidable error in a purity determination from the statistical standpoint would seem to result from two independent sources. There are (1) the errors connected with the identification and weighing of the impurities and (2) the error of random sampling.

The error of weighing is itself very complex, made up of errors from such sources as the sensitivity and reliability of the balance, errors in the weights used, fluctuations in the atmospheric humidity, and the accidental loss of material. The resultant of these and doubtless other factors will influence the observed weight of any particular sample. Repeated weighing of the same sample will give values that in all probability will be distributed normally about a mean with a standard deviation that will vary with different material. With carefully standardized methods it should be possible to determine a satisfactory average standard deviation for each type of seed.

In the absence of definite information regarding the magnitude of the standard deviation due to errors other than that of random sampling, it will be necessary to represent it by a symbol. Let σ_w stand for the standard deviation of weighing, etc., expressed as a percentage of the sample. If it is assumed that the impurities will be made up of discrete particles of approximately the same weight as the individual seeds being tested, the error of random sampling may be predicted with some degree of accuracy. This assumption does not seem unreasonable, since sieving and winnowing may be expected to remove all particles that differ materially from the seeds in either size or specific gravity. Under these conditions the variance in percentage is the product of the percentage of pure seed and the percentage of impurities divided by the estimated number of seeds in the sample.

The total standard deviation of the percentage of pure seed is the square root of σ_w^2 plus the variance due to random sampling.

Let N equal the number of seeds in the sample, p the percentage of pure seed, q the percentage of impurities, σ_w the standard deviation of weighing expressed as a percentage, σ the standard deviation of percentage of pure seed—

$$\text{then } \sigma = \sqrt{pq/n + \sigma_w^2} \qquad (5)$$

It would be of interest to compare the official tolerance with a tolerance based on the standard deviation. This can not be done accurately without a knowledge of the nature and size of the standard deviation of weighing. It seems probable, however, that this component will be relatively unimportant, and it may be of some interest to see what degree of certainty the official tolerance affords for samples of different sizes and percentages of pure seed, considering errors of random sampling only. For seed that is 99 per cent pure the official procedure provides a degree of certainty that ranges from 0.02 for samples that contain 2,000 seeds to 0.00001 for samples of 11,000 seeds. At 95 per cent the corresponding range is from 0.007 to 0.000000003, and at 85 per cent the range is from 0.00003 to something less than 10^{-10}.

These rather absurd probabilities for the lower percentages and larger samples will be somewhat reduced when errors of weighing are taken into consideration, but this will not alter the fact that the official tolerance gives a much higher degree of certainty for low percentages of purity than it does for high percentages.

TOLERANCE IN GERMINATION TESTS

From the standpoint of statistical treatment tolerance in germination tests appears to be a much simpler problem than purity tolerance. In germination tests there are two sources of error, (1) errors in the technic of germination and (2) errors of random sampling; but germination tests afford a rather definite measure of both types of error. With standardized conditions of germination, errors in technic should be small except for an occasional mistake. Fortunately, the standard procedure, that calls for the separate testing of four lots of 100 seeds each, should make it possible to detect gross errors except in those instances where the error is the same for all of the four lots. It would seem to be correct procedure to reject any test that shows a variation too large to be explained as an extreme fluctuation of sampling.

The expected variance of percentages for lots of 100 is p times q divided by 100. It will be remembered that p is the percentage of live seeds and q the percentage of dead seeds. The standard deviation is, of course, the square root of the variance.

There is some question regarding the most exact method for the rejection of aberrant percentages, but since the question does not involve a matter of justice but is to determine whether the test should be repeated, a simple method would seem to be adequate.

A very satisfactory method can be based on the magnitude of the difference between the percentage of germination of the aberrant 100 seeds and the percentage of germination of the other 300 measured by the standard deviation of the difference. The method can be made as conservative as desired by varying the degree of certainty. Since any one of the four members may be the outstanding percentage, the values given in the probability table should be increased accordingly, and a degree of certainty of 1 in 100 would be represented by 3.02 times the standard deviation instead of 2.33.

As an example of this method for rejecting aberrant germination tests, assume four tests each of 100 seeds giving germination percentages of $a=96$, $b=98$, $c=97$, and $d=87$. Can the test with 87 per cent be considered an extreme fluctuation? The degree of certainty is taken as 3.02, or 1 in 100.

Let M=mean percentage of germination of a, b, c, and $d=94.5$;
M_{a-c}=mean percentage of germination of a, b, and $c=97$;
M_d=percentage of germination of $d=87$;
$\Delta=M_{a-c}-M_d=10$;
N=number of seeds in a, b, and $c=300$;
N_d=number of seeds in $d=100$;
p=percentage of live seeds in $M=94.5$;
q=percentage of dead seeds in $M=5.5$;
σ_{a-c}^2=variance of $M_{a-c}=pq/N=94.5\times 5.5/300=1.73$;
σ_d^2=variance of $M_d=pq/N_d=94.5\times 5.5/100=5.20$;
$\sigma_\Delta=\sqrt{\sigma_{a-c}^2+\sigma_d^2}=\sqrt{1.73+5.20}=2.63$;
$\Delta/\sigma_\Delta=10/2.63=3.8$.

Since Δ/σ_Δ exceeds 3.02, the test should be rejected.

It would seem that the same method is applicable when the tests are made in duplicate only.

When the test consists of but two lots and the results are inconsistent, there is no way of telling which is erroneous, and a retest is necessary.

The retest may fall between the two previous determinations, bringing all four percentages within the range of random samples. In this case it would seem that the mean of all four should be taken as the best estimate of the true germination. On the other hand, the retest may demonstrate that one of the original percentages was erroneous, and then the mean should be based on the three remaining tests.

It would seem that the tolerance to be allowed in connection with the mean percentage of accepted germination tests should be determined by the standard deviation of that percentage. As previously explained, this is found by multiplying the percentage of live seeds by the percentage of dead ones, dividing by the number, and extracting the square root. (See formula 4, p. 10.) With the aid of the slide rule the operation should take less than 10 seconds, or, if tables are available, even less. To fix the tolerance, the standard deviation should be multiplied by the degree of certainty expressed in terms of the standard deviation.

The tolerances given in the official rules indicate a high degree of certainty. The lowest is that for percentages of 70. The official tolerance of 8 seeds at 70 per cent if based on 400 seeds is 3.49 times the standard deviation, or a probability of 1 in 4,000. At percentages above 95 the probability is too remote to be calculated from existing tables. If based on 200 seeds, the probabilities range from 1 in 147 to 1 in 435 for the lower percentages and 1 in some millions when percentages are above 95.

Perhaps there should be a word of caution against using the standard deviation of very high percentages in fixing tolerance. Unless based on very large numbers the *pq* standard deviation will be less than the true value. It should be a safe rule to restrict the application of the *npq* formula to tests where there are at least 10 dead seeds.

NOXIOUS WEEDS

Distributions that follow the Poisson series are rather uncommon in biological work, but when any particular kind of seed is present in very small quantities, the number of such seeds in successive samples from the same bulk may be expected to follow this distribution. As an example of one of the problems that the seed analyst meets, take dodder in clover seeds and assume that the maximum quantity of dodder allowed is represented by 10 dodder seeds in 50 grams of clover seed. Now, suppose that on analysis a given sample of 50 grams shows 15 dodder seeds. Is it safe to assert that the lot of seed from which the sample was drawn has in excess of 10 dodder seeds per 50 grams? As in all such cases, the first thing to decide is what is meant by safe, for this can never be certain. If this point is settled by saying that to be safe the answer in a series of similar decisions, on the average, must be right at least 99 times in 100, and if the method of sampling also is satisfactory, the work can then proceed. It is now a simple matter to answer the original question—reference to a table of Poisson distributions shows at once that from a bulk having 10 dodder seeds per 50 grams, samples with 15 or more should occur nearly five times in 100.

In order to keep the chosen degree of certainty it is not safe to reject any lot on the basis of a single sample unless this shows 19 or more dodder seeds. Presumably the procedure would be to repeat the test. This would certainly be necessary if errors on the side of too great leniency are to be avoided. Except for the possibility of detecting gross errors in sampling, the taking of two samples is exactly the same as a single sample of double the size. Suppose the second sample gave 17 dodder seeds. Obviously there is no discordance between the two samples and they may be combined. This gives 32 dodder seeds in 100 grams, where the limit is 20. Under a mean of 20 the table shows that an excess of 12 or more should occur only eight times in one thousand, and the lot may be rejected with the assurance that injustice would be done less often than once in one hundred times.

Once the degree of certainty is adopted, the limits of tolerance can be read directly from the published tables. In fixing a degree of certainty it should be kept in mind that a probability of 0.01, or one mistaken finding in 100, does not mean a mistake in 1 per cent of the tests made, but only in 1 per cent of the cases similar to that under investigation. The probability 0.01 has been taken arbitrarily for purposes of illustration. Perhaps it is too rigid. If not and the limits of tolerance are to be fixed between other than wide limits it will be necessary to test rather large samples.

In determining the size of the sample it will be well to remember that errors of random sampling are independent of the percentage of noxious seeds present and are governed entirely by the number of noxious seeds in the sample. The size of the sample should be so chosen that the stated number of seeds in the sample will give the desired accuracy. The published tables do not give the distribution of numbers above 30, but for numbers larger than 30 the distribution of departures above and below the mean are practically alike, and resort may be had to the old measure of accuracy, the standard deviation, and the use of the probability tables.

In a Poisson series the standard deviation equals the square root of the mean. Thus, if a sample were taken of such size that the stated or permitted number of seeds was 100 the standard deviation would be 10. A range of 2.3 times the standard deviation is 23 seeds above and 23 seeds below 100, and the tolerance would be 77 to 123 if the degree of certainty is to be maintained at 1 in 100. If the degree of certainty is lowered to say 0.04, or 1 in 25, the tolerance would be reduced to 17 instead of 23 seeds.

There would seem to be three factors or variables involved in determinations of noxious weeds. There is (1) the number of noxious seeds in the sample, (2) the tolerance or number of seeds to be allowed in excess of the stated number, and (3) the degree of certainty. These three factors are interrelated, and if any two of them are fixed the third is easily determined. In tests involving as many as 30 noxious seeds no tables are needed.

If the tolerance is expressed as a percentage of the number of noxious seeds and the standard of certainty in terms of the standard deviation, the relationship is given by the equation: Tolerance equals the degree of certainty divided by the square root of the number. If tolerance is expressed as the number of seeds, the equation is: Tolerance equals the square root of the number times the degree of certainty. As an illustration, suppose it is desired to know the tolerance in excess that must be allowed in tests that should contain 50 noxious seeds as the stated number, the degree of certainty being 2.33, which corresponds to 1 mistake in 100. The tolerance then is 2.33 times the square root of 50, or 16.5, say 17, seeds. If with this same lot it is desired to know how large a sample would need to be tested to fix the tolerance at 20 per cent, the square root of the number of seeds is ascertained by dividing 2.33 by 0.2. This equals 11.7, and since this is the square root of the number, the square of it, or 136, is the number. That is, the sample would need to be of a size that would have 136 noxious seeds if these were present in the stated proportion.

These relations are more concisely stated as follows:

Let C=standard of certainty expressed in terms of the standard deviation,
M=number of noxious seeds in sample,
T=tolerance (expressed in number of seeds),
$T\%$=tolerance expressed as a ratio$=T/M$;

then $T=C\times\sqrt{M}$
$T\%=C/\sqrt{M}$
$M=(C/T\%)^2$
$C=T\%\times\sqrt{M}$, or $T/\sqrt{M}$

A somewhat related problem that may arise involving the Poisson series is to determine the probability that the difference between two samples supposed to have come from the same bulk may be due to errors of random sampling. If the samples contain 30 or more noxious seeds a normal distribution may be assumed; and, since the standard deviation is the square root of the number, the standard deviation of the difference will be the square root of twice the mean number. The difference divided by this standard deviation gives the value needed to enter the normal probability table. If the mean number is much less than 30, however, this procedure will give a standard deviation that is too small. It would seem that the proper procedure is to combine the probabilities given in the Poisson tables. A close approximation to the correct result is to take four times the product of the two probabilities.

Suppose two samples gave 4 and 16 seeds. The assumed mean is 10, the table gives 0.029 as the probability of a minus departure of 6 and 0.049 as the probability of a plus departure of 6; $0.029\times0.049\times4=0.0057$; that is, there is 1 chance in 175 that the two samples came from the same population.

CONCLUSION

It is realized that the foregoing are but a few of the problems confronting the seed analyst that call for statistical treatment. The examples have been chosen as representative, and the aim has been to show that methods involving neither complicated mathematics nor laborious calculations may add precision to the decisions that the seed analyst is called upon to make.

If a suggestion is permitted, it would be that in seed testing the adoption of standard and uniform methods for the statistical treatment of results is of the same importance as standard methods of testing. Both are needed as a protection to the individual worker.

ORGANIZATION OF THE UNITED STATES DEPARTMENT OF AGRICULTURE

October 11, 1929

Secretary of Agriculture	ARTHUR M. HYDE.
Assistant Secretary	R. W. DUNLAP.
Director of Scientific Work	A. F. WOODS.
Director of Regulatory Work	WALTER G. CAMPBELL.
Director of Extension Work	C. W. WARBURTON.
Director of Personnel and Business Administration.	W. W. STOCKBERGER.
Director of Information	M. S. EISENHOWER.
Solicitor	R. W. WILLIAMS.
Weather Bureau	CHARLES F. MARVIN, *Chief.*
Bureau of Animal Industry	JOHN R. MOHLER, *Chief.*
Bureau of Dairy Industry	O. E. REED, *Chief.*
Bureau of Plant Industry	WILLIAM A. TAYLOR, *Chief.*
Forest Service	R. Y. STUART, *Chief.*
Bureau of Chemistry and Soils	H. G. KNIGHT, *Chief.*
Bureau of Entomology	C. L. MARLATT, *Chief.*
Bureau of Biological Survey	PAUL G. REDINGTON, *Chief.*
Bureau of Public Roads	THOMAS H. MACDONALD, *Chief.*
Bureau of Agricultural Economics	NILS A. OLSEN, *Chief.*
Bureau of Home Economics	LOUISE STANLEY, *Chief.*
Plant Quarantine and Control Administration.	C. L. MARLATT, *Chief.*
Grain Futures Administration	J. W. T. DUVEL, *Chief.*
Food, Drug, and Insecticide Administration	WALTER G. CAMPBELL, *Director of Regulatory Work, in Charge.*
Office of Experiment Stations	E. W. ALLEN, *Chief.*
Office of Cooperative Extension Work	C. B. SMITH, *Chief.*
Library	CLARIBEL R. BARNETT, *Librarian.*

This circular is a contribution from

Bureau of Plant Industry	WILLIAM A. TAYLOR, *Chief.*
Seed Laboratory	EDGAR BROWN, *Principal Botanist, in Charge.*
Biophysical Laboratory	G. N. COLLINS, *Principal Botanist, in Charge.*

U. S. GOVERNMENT PRINTING OFFICE: 1929

For sale by the Superintendent of Documents, Washington, D. C. - - - - - Price 5 cents